15400 Calhoun Drive, Suite 400
Rockville, Maryland, 20855
(301) 294-5200
http://www.i-a-i.com

Intelligent Automation Incorporated

Enhancements for a Dynamic Data Warehousing and Mining System for Large-scale HSCB Data

Monthly Report No. 5

Reporting Period: July 20, 2016 – Aug 19, 2016

Contract No. N00014-16-P-3014

Sponsored by
ONR, Arlington VA
C⊙TR/TP⊙C: Dr. Rebecca Goolsby

Prepared by
Onur Savas, Ph.D.

DISTRIBUTION A

Approved for public release; distribution is unlimited.

Monthly Report No. 5

Enhancements for a Dynamic Data Warehousing and Mining System Large-Scale HSCB Data

Submitted in accordance with requirements of
Contract #N00014-16-P-3014

Performance period: July 20, 2016 to Aug 19, 2016
(PI: Dr. Onur Savas, 301.294.4241, osavas@i-a-i.com)

1 Work Performed within This Reporting Period

In this reporting period, we performed the following tasks.

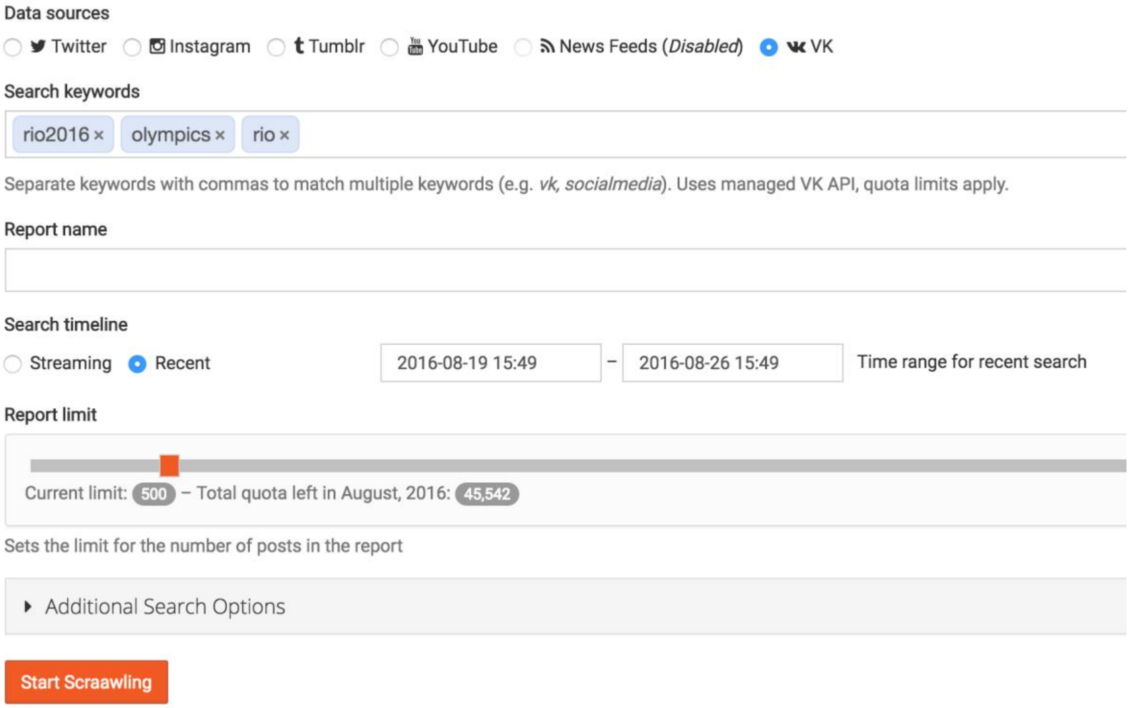
- **Developed VK Data Collection and Basic Statistics Computation Capabilities.** We have developed a capability to retrieve VK posts based on keywords and geospatial locations. We have also developed a capability to compute basic statistics of the VK posts that include computation of top users, top words, top URLs, and top attachments along with top popular media. These capabilities are integrated into Scraawl.

- **Released Scraawl 1.18.0**

1.1 VK Data Collection

VK is the largest European online social networking service. It is available in several languages, but is especially popular among Russian-speaking users [1]. Similar to other social networks, VK allows users to message each other publicly or privately, to create groups, public pages and events, share and tag images, audio and video, and to play browser-based games [1].

In this reporting period, we have matured VK crawling capabilities. In particular, we have developed a capability to retrieve VK posts based on keywords. Current search grammar includes combining a set of keywords by AND'ing or OR'ing them. The search can be performed as streaming or a 1-week historical, and can be combined with a geolocation search. The search is integrated as part of Scraawl, and the UI is shown in Figure 1. Similar to other social network searches, the keywords or phrases can be entered separately, a report name can be given to the search, and either a streaming or a historical search can be chosen. In addition, a map is interactively used (not shown) to restrict the search to a geospatial region.



Data sources

○ Twitter ○ Instagram ○ Tumblr ○ YouTube ○ News Feeds (*Disabled*) ● VK

Search keywords

rio2016 × olympics × rio ×

Separate keywords with commas to match multiple keywords (e.g. *vk, socialmedia*). Uses managed VK API, quota limits apply.

Report name

Search timeline

○ Streaming ● Recent 2016-08-19 15:49 – 2016-08-26 15:49 Time range for recent search

Report limit

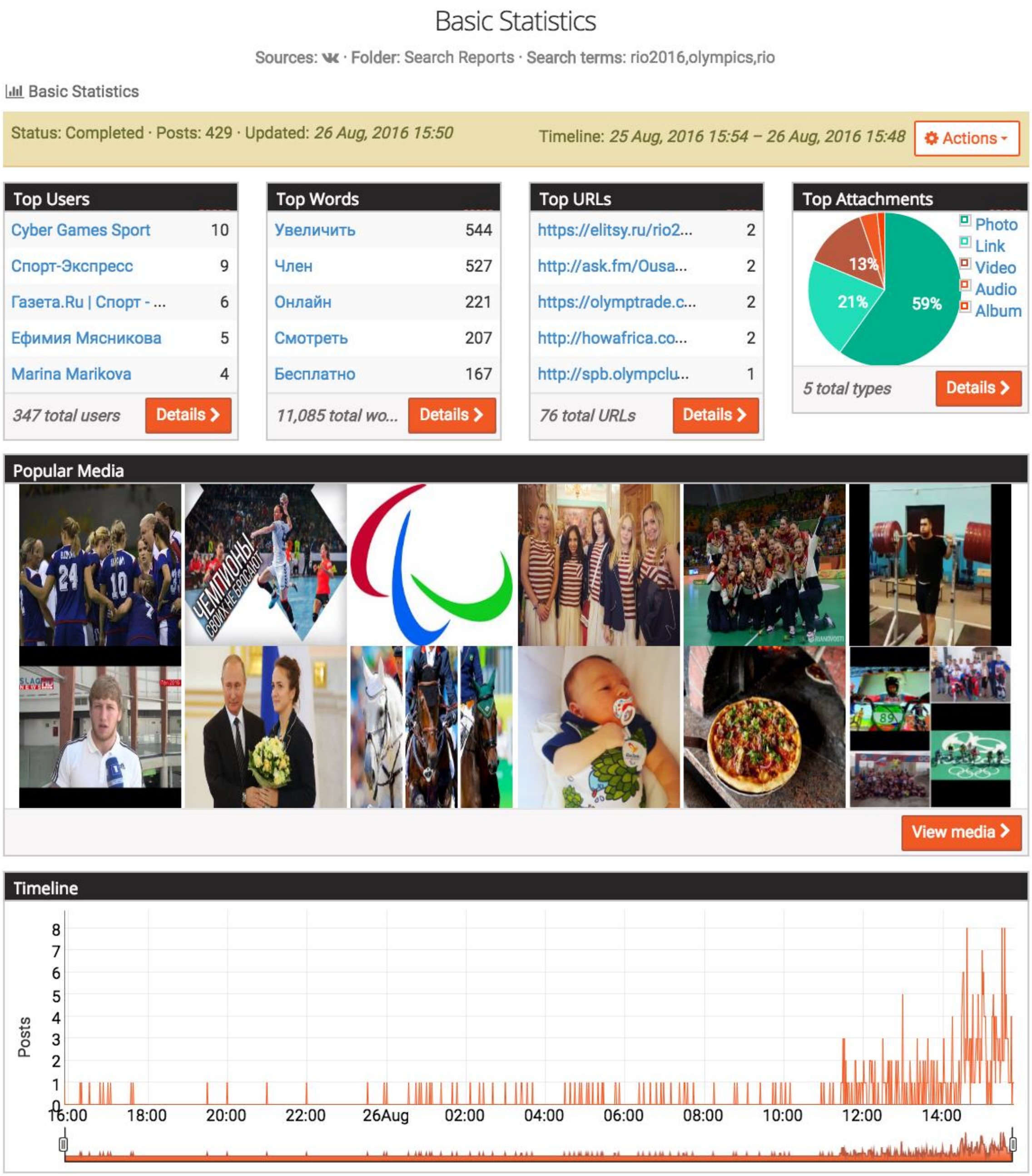
Current limit: 500 – Total quota left in August, 2016: 45,542

Sets the limit for the number of posts in the report

Figure 1: Representative VK Search Screen.

1.2 VK Basic Statistics

We have also developed capabilities to compute basic statistics of the VK posts that were collected using the interface of Figure 1. In particular, top users, top words, top URLs, and top attachments along with top popular media is computed and presented to the user in one screen as shown in Figure 2. The timeline of the posts (not shown) is also presented in the same screen. Each top statistics (e.g., top words) can be further drilled down to show additional information.



Basic Statistics

Sources: · Folder: Search Reports · Search terms: rio2016,olympics,rio

Basic Statistics

Status: Completed · Posts: 429 · Updated: *26 Aug, 2016 15:50*

Timeline: *25 Aug, 2016 15:54 – 26 Aug, 2016 15:48*

Top Users	
Cyber Games Sport	10
Спорт-Экспресс	9
Газета.Ru | Спорт - ...	6
Ефимия Мясникова	5
Marina Marikova	4
347 total users	Details

Top Words	
Увеличить	544
Член	527
Онлайн	221
Смотреть	207
Бесплатно	167
11,085 total wo...	Details

Top URLs	
https://elitsy.ru/rio2...	2
http://ask.fm/Ousa...	2
https://olymptrade.c...	2
http://howafrica.co...	2
http://spb.olympclu...	1
76 total URLs	Details

Top Attachments

5 total types

Figure 2: Representative VK Basic Statistics View.

2 Current Problems

None.

3 Work to be Performed in the Next Reporting Period

In the next report period, we will focus on the following tasks:

- We will deliver Scraawl 2.0.

References

[1] https://en.wikipedia.org/wiki/VK_(social_networking).